AD-A144 522   INFORMATION-THEORETIC NON-PARAMETRIC UNIMODAL DENSITY   1/1
               ESTIMATION(U) TEXAS UNIV AT AUSTIN CENTER FOR
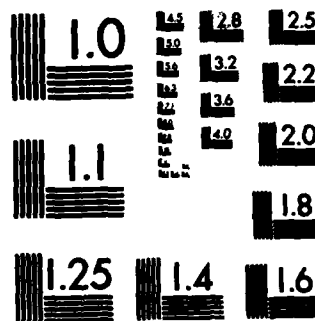   CYBERNETIC STUDIES   P BROCKETT ET AL. FEB 84 CCS-RR-467
UNCLASSIFIED   N00014-81-K-0145   F/G 9/4   NL

END

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

(12)

INFORMATION-THEORETIC NON-PARAMETRIC UNIMODAL DENSITY ESTIMATION

by

P. Brockett
A. Charnes
K. Paick

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

DTIC
ELECTE
AUG 2 1 1984

B

84 08 13 006

AD-A144 522

Research Report CCS-467 ✓

INFORMATION-THEORETIC NON-PARAMETRIC
UNIMODAL DENSITY ESTIMATION

by

P. Brockett
A. Charnes
K. Paick

February 1984

DTIC
ELECTE
AUG 2 1 1984

B

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Graduate School of Business 4.138
The University of Texas at Austin
Austin, Texas 78712
(512) 471-1821

$'a-$

INFORMATION-THEORETIC NON-PARAMETRIC
UNIMODAL DENSITY ESTIMATION

by

P. Brockett, A. Charnes, K. Paick

Abstract

→ This document

We present an information-theoretic method for nonparametric density estimation which guarantees that the resulting density is unimodal. The method inputs data in the form of moment or quantile information and consequently can handle both data derived and non-data derived information. In the non-data derived situation it yields a method for obtaining unimodal Bayesian prior distribution.

Key Words

Information Theory
M.D.I. Density Estimation
Maximum Entropy
Unimodality
Non-Parametric Density Estimation

I. INTRODUCTION

In many problems encountered in engineering and signal processing, it is useful to estimate the probability density function corresponding to some random phenomenon under study. Often it is known that the density is unimodal, and the first few cumulants can effectively be estimated from the available data. However, the precise parametric formula for the generating density cannot be determined from physical considerations alone and often might not match any of the commonly assumed densities, even if a parametric model was appropriate. Accordingly, the usual technique in applied research is to exercise a procedure of nonparametric density estimation and then numerically obtain the estimated density. There is usually no closed form solution to such procedures (e.g., window estimates or kernel estimates), although numerical computations such as likelihood ratio detection algorithms can be implemented using these techniques. Unfortunately, the resulting nonparametric density estimate obtained by smoothing the empirical histogram usually does not exhibit unimodality even when it is known that the true density is unimodal. The tail behavior of these density estimates may not be monotonic. It is, however, exactly these tails which are frequently the major object of interest.

In this paper we present a new method for unimodal nonparametric density estimation which is not based upon smoothing the empirical histogram per se. Instead, it uses a method of Kemperman's [1] for transforming moment problems, and couples this with an information-theoretic generalization of Laplace's famous "principle of insufficient reason" to obtain a unimodal density estimate. The resulting density

1

estimate is rendered in closed analytic form and can easily be worked with numerically, e.g., for optimal (likelihood ratio or Neyman-Pearson) signal detection in noise.

It should be remarked that our method is also applicable to the situation of prior information which is not necessarily data derived and can be used for assessing a unimodal prior distribution for subsequent Bayesian analysis. This topic will be pursued in a separate paper [2].

In section II we present the information-theoretic estimation procedure that we have employed. In section III we present the method used to transform the problem of unimodal density estimation to an estimation problem involving an auxiliary variable. In section IV we present the actual nonparametric estimation procedure which ensures a unimodal density estimate and which exhibits certain pertinent characteristics desired. In the final section, we give some numerical results using both real and simulated data.

II. MAXIMUM ENTROPY AND MDI DENSITY ESTIMATION

The concept of statistical information and density estimation for numerical data is of paramount importance in statistics, economics, engineering, signal detection, and other fields. Wiener [3] remarked quite early, in 1948, that the Shannon measure of information from statistical communication theory could eventually replace Fisher's concept of statistical information for a sample. For example, using a measure of informational distance between two measures first developed by Kullback and Leibler [4] in 1951, following the work of Khinchin, it has been shown how to estimate the order in an autoregressive time series model, how to estimate the number of factors in a factor analysis model, and how to analyze contingency tables (cf. Akaike [5,6] and Gokhale and

2

Kullback [7]). Minimizing this statistical "distance" subject to the given constraints is called "Khinchin-Kullback-Leibler" (K^2L) estimation, or "minimum discrimination information" (MDI) estimation in the literature. Mathematically the problem is to pick that density function f which is "as close as possible" to some given function g, and which satisfies certain given moment constraints, e.g.,

$$\min_f \int f(x) \, \ell n \left(\frac{f(x)}{g(x)}\right) \lambda(dx),$$

subject to

$$\int a_0(x) f(x) \lambda(dx) = 1 = \theta_0$$

$$\int a_1(x) f(x) \lambda(dx) = \theta_1 \qquad\qquad (2.1)$$

$$\cdot \qquad \cdot$$

$$\cdot \qquad \cdot$$

$$\cdot \qquad \cdot$$

$$\int a_k(x) f(x) \lambda(dx) = \theta_k \, .$$

Here λ is some dominating measure for f and g (usually Lebesgue measure in the continuous case, or counting measure in the discrete case), $\theta_1, \ldots, \theta_k$ are the given moment values for the moment functions $a_1, \ldots, a_k$, and $a_0(x) \equiv 1$. The moment functions $a_i(x)$ may be used to generate moment or cumulant constraints, e.g., when a_i is a polynomial, or may generate centile constraints, e.g., when a_i is an indicator function for a terminal interval.

In many applications there is no a priori choice of a distribution g to serve as a goal density in (2.1). In this case we express our ignorance by choosing all x values to be equally likely, i.e., $g(x) \equiv 1$. In this case the MDI or K^2L objective functional is of the form

$$\int f(x) \ell n f(x) \, \lambda(dx) = -\int f(x) \ell n \frac{1}{f(x)} \, \lambda(dx) \qquad \cdot$$

3

This is precisely minus the entropy of the density, and the MDI problem becomes a maximum entropy (ME) problem. The ME criterion can be thought of as taking the most "uncertain" distribution possible subject to the given constraints. Accordingly, this principle of maximum entropy may be construed as a new and important extension of the famous Laplace "principle of insufficient reason," which postulates a uniform distribution in situations in which nothing is known about the variable in question. (See Guiasu [8] for more motivation and explanation of these results.) Here the ME distribution is as close to uniform as possible, subject to the given informational constraints.

The minimization of (2.1) subject to the given constraints is easily carried out by Lagrange multipliers. The short derivation given below can essentially be found in Guiasu [8].

Introducing a Lagrange multiplier for each constraint in (2.1) and, changing from a minimization to a maximization, we wish to maximize

$$L = \int f(x) \, \ln \frac{g(x)}{f(x)} \, \lambda(dx) - \alpha_0 \left[\theta_0 - \int a_0(x) f(x) \lambda(dx) \right]$$

$$+ \cdots + \alpha_k \left[\theta_k - \int a_k(x) f(x) \lambda(dx) \right]$$

or, equivalently,

$$L - \sum \alpha_k \theta_k = \int f(x) \left\{ \ln \left[\frac{g(x)}{f(x)} \right] - \alpha_0 - \alpha_1 a_1(x) - \alpha_2 a_2(x) - \cdots - \alpha_k a_k(x) \right\} \lambda(dx)$$

$$= \int f(x) \, \ln \left[\frac{g(x) \exp\left\{ -\sum_{i=0}^{k} \alpha_i a_i(x) \right\}}{f(x)} \right] \lambda(dx)$$

$$\leqslant \int f(x) \left[\frac{g(x) \exp\left\{ -\sum_{i=0}^{k} \alpha_i a_i(x) \right\}}{f(x)} - 1 \right] \lambda(dx)$$

4

The inequality follows since $\ln x \leqslant x-1$ with equality only at $x=1$. Thus, the inequality becomes an equality when

$$f(x) = g(x)\exp\left[-\sum_{i=0}^{k} \alpha_i a_i(x)\right] ,$$

and this becomes the maximizing density. We shall call (2.2) the MDI density (or the ME density if $g(x)\equiv 1$) subject to the constraints.

The numerical value of the constants α_i are found using the moment constraints. In Brockett, Charnes, and Cooper [9] or Charnes, Cooper, and Seiford [10], it is shown how to obtain the constants α_i as dual variables in an unconstrained convex programming problem. We shall discuss this fact further in the section concerning numerical computation, since in the dual formulation of the MDI problem the computation is easily accomplished using any of a number of existing nonlinear programming codes.

III. TRANSFORMING THE UNIMODAL DENSITY ESTIMATION PROBLEM

In this section we will show how to use the information that a density is unimodal in the nonparametric estimation problem. The technique to be used is borrowed from Professor J. H. B. Kemperman, whose 1971 paper [1] also gives more advanced and wide ranging moment transformation techniques.

A famous characterization of "zero" unimodal random variables (due to L. Shepp following the work of Khinchin) is the following. Suppose Y is unimodal with mode zero. Then $Y=U \cdot X$ where U and X are independent, and U is uniformly distributed over $[0,1]$. A proof of this result can be found, for example, in Feller (Ref. 11, p. 158). From the above result, it follows immediately that

5

$$Eh(Y) = Eh*(X) \quad ,$$

where

$$h*(x) = E\big[h(UX)|X=x\big] = \frac{1}{x} \int_0^x h(t)dt \quad .$$

Our technique for solving the problem (2.1) in the unimodal case may now be explained as follows. If Y is unimodal with mode m, then Y-m is zero ' unimodal. First transform the given moment constraints $\int a_i(x) f_Y(x) d\lambda(x) = \theta_i$ on the original variable Y to constraints of the form $\int a_i^*(x) f_X(x) d\lambda(x) = \theta_i$ on an auxiliary variable X, where

$$a_i^*(x) = \frac{1}{x} \int_0^x a_i(t+m)dt \quad .$$

If the mode is unknown, then a consistent estimator $\hat{m}$ may be used. (See Sager [12] for such a nonparametric mode estimator.) We then solve the transformed MDI problem involving the constrained estimation of f_X. Using the estimated X density we then transform back to obtain the estimate for Y. If X is estimated by $\hat{X}$, then Y is estimated by $m+U\cdot\hat{X}$, and consequently is unimodal by Khinchin's theorem. The details are given in the next section.

IV. OBTAINING THE ESTIMATED DENSITY

Decomposing the original variable Y via the Khinchin representation $Y-m=U\cdot X$, we can now transform the constraint set in (2.1) into constraints involving X. Namely, by Kemperman's technique,

$$\theta_i = E\big[a_i(Y)\big] = E\big[a_i(Y-m+m)\big] = E\big[a_i^*(X)\big] ,$$

where

$$a_i^*(x) = \frac{1}{x} \int_0^x a_i(t+m)dt = \frac{1}{x} \int_m^{x+m} a_i(t)dt \quad .$$

As an illustration, if an original constraint on Y is a raw moment, say

$$a_1(x) = x^k \quad ,$$

6

then it follows that

$$a_1^*(x) = \left[(x+m)^{k+1} - m^{k+1}\right]\big/(k+1)x$$

$$= \sum_{j=0}^{k} \binom{k+1}{j} x^j m^{k-j} \big/ (k+1).$$

Thus the corresponding constraint in the auxiliary variable X involves a sum of moments up to order k. Similarly, if the constraint upon Y is a probability or centile constraint we obtain, for example,

$$\pi = P[Y > \xi] = E\left[I_{[\xi,\infty)}(Y)\right] = E\left[I_{[\xi-m,\infty)}(Y-m)\right] = E[a^*(x)],$$

where

$$a^*(x) = \frac{1}{x} \int_0^x I_{[\xi-m,\infty)}(t)\,dt = \begin{cases} 0, & x \leq \xi-m \\[2mm] 1 - \dfrac{\xi-m}{x}, & x \geq \xi-m \end{cases}$$

Having transformed the moment constraints on Y into constraints on X, we then estimate the density for X via maximum entropy, viz., f_X is the solution to

$$\max \; -\int f_X(x) \, \ell n \, f_X(x) \, d\lambda(x)$$

subject to

$$\theta_o = 1 = \int f_X(x)\lambda(dx)$$

$$\theta_1 = \int a_1^*(x) f_X(x)\lambda(dx)$$

$$\vdots \qquad \vdots$$

$$\theta_k = \int a_k^*(x) f_X(x)\lambda(dx) \qquad .$$

7

Using the sample data we are able to estimate the parameters $\theta_1, \ldots, \theta_k$ for the desired moment functions $a_1(x), a_2(x), \ldots, a_k(x)$ by $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$. According to formula (2.2) the estimate for the density of X is

$$\hat{f}_X(x) = \exp\left\{-\sum_{i=0}^{k} \alpha_i a_i^*(x)\right\} \qquad . \qquad (4.2)$$

The numerical value for the constants α_i can be determined using the dual convex programming formulation outlined in Brockett, Charnes, and Cooper [9].

An additional advantage of the ME procedure described above is that since (4.2) defines a member of the exponential family of distributions, the usual results concerning statistical properties (such as the existence of sufficient statistics) are valid. Under mild conditions on the estimator $\hat{\theta}$ of θ, it also follows that the parameter estimates are asymptotically normally distributed with derivable covariance matrix (cf. Kullback [13]). This allows for confidence interval statements to be made concerning $\hat{f}_X$.

Now, since f_X is estimated via (4.2), we may transform back to obtain the distribution for Y. Using the relationship $Y-m=U \cdot X$ we have the estimated density function for Y:

$$f_Y(y) = f_{Y-m}(y-m) = \int_0^1 f_X\left(\frac{y-m}{u}\right) \frac{du}{u} = -\int_{\mathrm{sgn}(y-m)}^{y-m} f_X(x) \frac{dx}{x} \qquad . \quad (4.3)$$

So

$$\hat{f}_Y(y) = -\int_{\mathrm{sgn}(y-m)}^{y-m} \exp\left\{-\sum_{i=0}^{k} \alpha_i a_i^*(x)\right\} \frac{dx}{x}$$

$$
= \begin{cases} \int_{-\infty}^{y-m} \exp\left\{-\sum \alpha_i a_i^*(x)\right\} \dfrac{dx}{x} & \text{if } y \leq m \\[2em] \int_{y-m}^{\infty} \exp\left\{-\sum \alpha_i a_i(x)\right\} \dfrac{dx}{x} & \text{if } y \geq m \end{cases} \qquad . \qquad (4.4)
$$

The unimodality of this density estimate is apparent since $-yf_Y'(y+m)=f_X(y) \geq 0$.

In the next section we give some examples using the above techniques.

V. NUMERICAL RESULTS

In this section we present several examples of the various types of constraints (moment functions) and the resulting graphical display of the estimated density. In the simulations we shall also display the "true" density from which the data were obtained. Estimating the density for X via minimum MDI (1.2) can easily be done by the duality theory given in Charnes, Cooper, and Seiford [10].

The unimodal estimation parameters $\{\alpha_i\}$ in the density $f_X(x)$ are dual variables in the unconstrained convex programming problem (5.1), with $a_i^*(x)$ replacing $a_i(x)$ and $\lambda(dx)=dx$ in (2.1). According to the duality result of Charnes, Cooper, and Seiford [10], the problem (2.1) has an unconstrained dual problem.

$$
\max_{\alpha} \sum_{i=0}^{k} \alpha_i \theta_i - \int g(x) \exp\left[-\sum_{i=0}^{k} \alpha_i a_i^*(x)\right] dx \qquad . \qquad (5.1)
$$

If the density is to be estimated over the entire interval $(-\infty, \infty)$, then some algebraic moment constraint must be included as one of the constraining equations. For the numerical examples given in this section we have used the expected value of the corresponding moment constraint on

9

the auxiliary variable X which is $E\sqrt{|y|}=\frac{2}{3}\int\sqrt{|x|}f_X(x)dx$.

A further result which should be noted is that if one wishes to impose a continuity constraint upon the estimated density f_Y at the mode m, then this continuity constraint of Y translates into another moment constraint upon the auxiliary variable X of the form

$$0 = \int a_{k+1}^{*}(x)\ f_X(x)dx \quad ,$$

where $a_{k+1}^{*}(x)=1/x$ (there is no corresponding $a_{k+1}(x)$ constraint on Y, but from (4.3) we see the constraint (5.2) amounts to $f_Y(m-o)=f_Y(m+o)$).

Three different goal densities were chosen for illustration:

$$g_1(x) = 1 \quad ,$$

$$g_2(x) = \begin{cases} \exp\left[\alpha-(|x|-\delta)^2-\left(\frac{1}{|x|}-\frac{1}{\delta}\right)^2\right] & \text{for } |x|\leqslant\delta \\ \\ 1 & \text{for } |x|\geqslant\delta \end{cases} \quad ,$$

$$g_3(x) = \frac{x^2}{\sqrt{2\pi}\ \sigma^3}\ \exp\left[\frac{-x^2}{2\sigma^2}\right] \quad .$$

The goal density $g_1(x)=1$ corresponds to maximum entropy estimation for f_X. To impose smoothness on the estimated density f_Y near the mode m, the goal density $g_2(x)$ was used. This goal density behaves like the constant 1 outside $|x|<\delta$, and dips smoothly to zero as $|x|\to0$, i.e. the goal density approximates the ME procedure given by $g_1(x)$, but constrains the estimated density f_Y to be smooth around the mode. The goal density $g_3(x)$ corresponds to the f_X density which would result from f_Y being normally distributed. Hence this goal density gives the "close to normality subject to constraints" interpretation for the estimated density f_Y.

10

Three examples of predictive distributions are presented. These are simulated observations from a normal distribution, a Cauchy distribution, and some real data from acoustical returns generated by biological undersea noise in ambient background noise.

Normal Distribution

The following constraints were imposed:

(1) Y is unimodal with the most likely value for m as 5.

(2) $Pr[-\infty \leq Y \leq 5] = 0.5$.

(3) $Pr[4 \leq Y \leq 6] = 0.6826$.

Figure 1 shows the estimated distribution for each of the goal densities. Note that, when the data is truly normal, reasonable estimates are obtained using only the three pieces of information above. Our procedure can also be used even if there are no data, but just auxiliary information (cf. Brockett, Charnes, and Paick [2]).

Cauchy Distribution

To determine the performance of this method when the moment constraints have been derived from data, we generated 1000 Cauchy random numbers using the congruential generator RANF in the standard FORTRAN library. The following moment constraints were prescribed using estimates derived from the data.

(1) Y is unimodal with possible value between -400 and 400.

(2) The most likely value for m is 0.

(3) The distribution of Y is symmetric about 0.

(4) $P[-3.5325 \leq Y \leq 3.5325] = 0.8$.

(5) Expected value of $\sqrt{|Y|} = 1.427552$.

Figure 2 shows the results of the calculation using the above information for each of the goal densities.[*]

Distribution of Underwater Acoustic Returns

The ambient noise was recorded in a shallow water coastal region which was dominated by acoustic energy from snapping shrimp. 30,000 recorded samples were used to obtain a distribution for these underwater acoustic returns. The following information was obtained from the data:

1. Y is unimodal with possible values between -1024 and 767.

2. The most likely value for m is 0.

3. The following probabilities are obtained from the data:

$$Pr[-1024 \leq Y \leq 0] = 0.50$$

$$Pr[-1024 \leq Y \leq -71] = 0.05$$

$$Pr[-20 \leq Y \leq 0] = 0.25$$

$$Pr[0 \leq Y \leq 20] = 0.25$$

$$Pr[0 \leq Y \leq 71] = 0.45$$

4. Expected value of $\sqrt{|Y|}$ = 5.03, as estimated from the data. Figure 3 shows the predictive distribution for each of the goal densities based on this information.

---

[*]$(Range/4)^2$ is adopted as a variance of the "normal" goal density $g_3(x)$ for illustrative purposes. Although the variance of the Cauchy distribution does not exist, the normal shape can still be a permissible goal density.

REFERENCES

[1] J. H. Kemperman, "Moment Problems with Convexity Conditions I", in *Optimizing Methods in Statistics*. New York: Academic Press, 1971.

[2] P. L. Brockett, A. Charnes, and K. H. Paick, "A method for obtaining a unimodal prior distribution density", Report CCS 473, Center for Cybernetic Studies, The University of Texas at Austin, 1983.

[3] N. Wiener, *Cybernetics*. New York: John Wiley and Sons, Inc., 1948.

[4] S. Kullback and R. A. Leibler, "On information and sufficiency", Ann. Math. Stat., vol. 22, pp. 76-86, 1951.

[5] H. Akaike, "An extension of the model of maximum likelihood and Stein's problem", Ann. Inst. Statist. Math., vol. 29, part A, pp. 153-164, 1977.

[6] H. Akaike, "Information theory and an extension of the maximum likelihood principle", Proc. of the 2nd Int. Symp. Inform. Theory, 1972.

[7] D. V. Gokhale and S. Kullback, *The Information in Contingency Tables*. New York: Marcel Dekker, Inc., 1978.

[8] S. Guiasu, *Information Theory with Applications*. Great Britain: Mc-Graw-Hill Book Co., Inc., 1977.

[9] P. L. Brockett, A. Charnes, and W. W. Cooper, "MDI estimation via unconstrained convex programming", Commun. Stat. Simula. and Computa., vol. B9(3), p. 223, 1980.

[10] A. Charnes, W. W. Cooper, and L. Seiford, "Extremal principles and optimation qualities for Kinchine-Kullback-Leibler estimation", Math. Operationsforsch. Statist. Ser. Optimization, vol. 9, no. 1, pp. 21-29, 1978.

[11] W. Feller, *An Introduction to Probability Theory and Its Application*, Vol. 2. New York: John Wiley and Sons, 1971

[12] T. Sager, "Consistency in non-parametric estimation of the mode", Ann. of Stats., vol. 3, pp. 698-706, 1975.

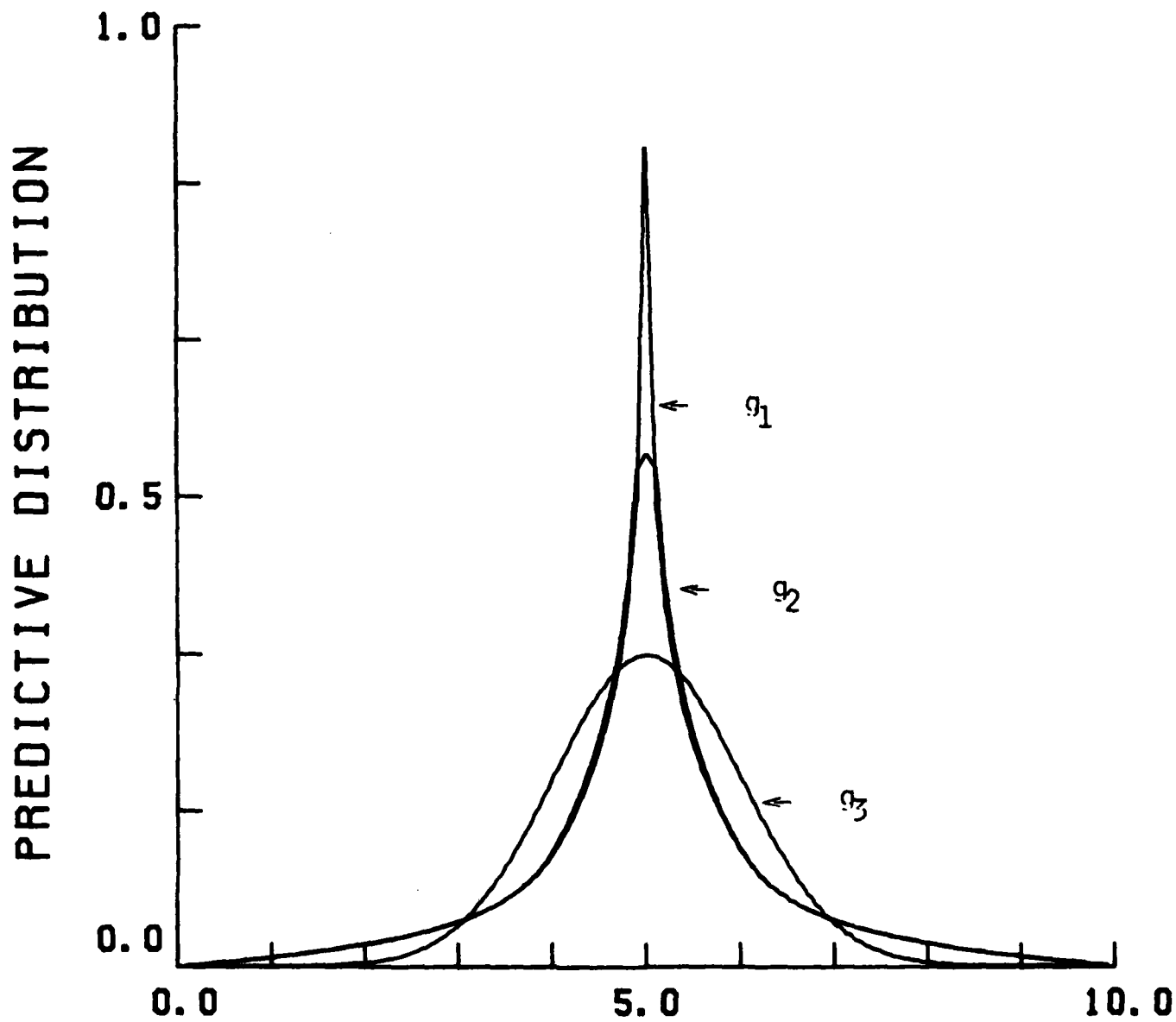[13] S. Kullback, *Information Theory and Statistics*. New York: John Wiley and Sons, 1959.
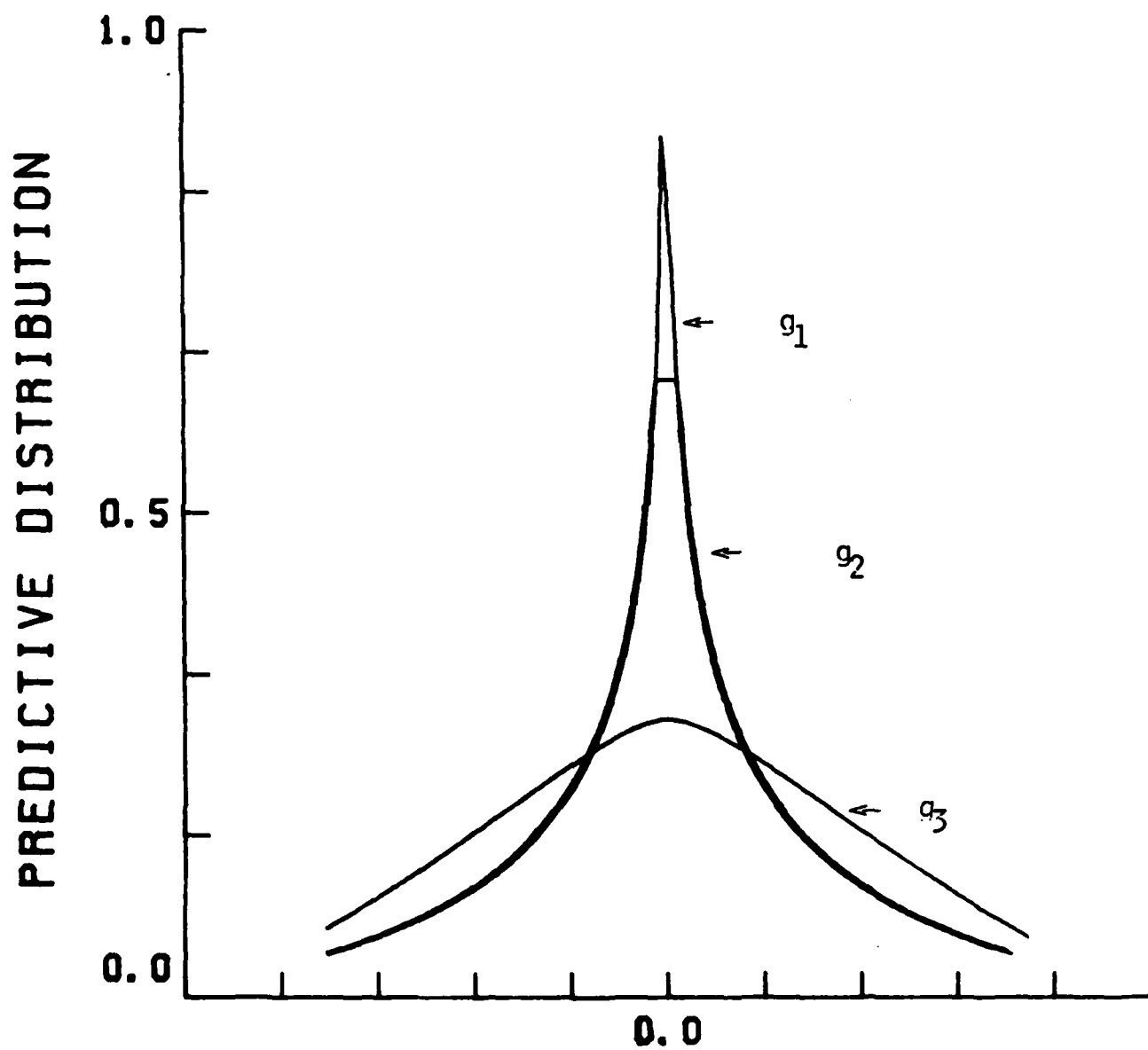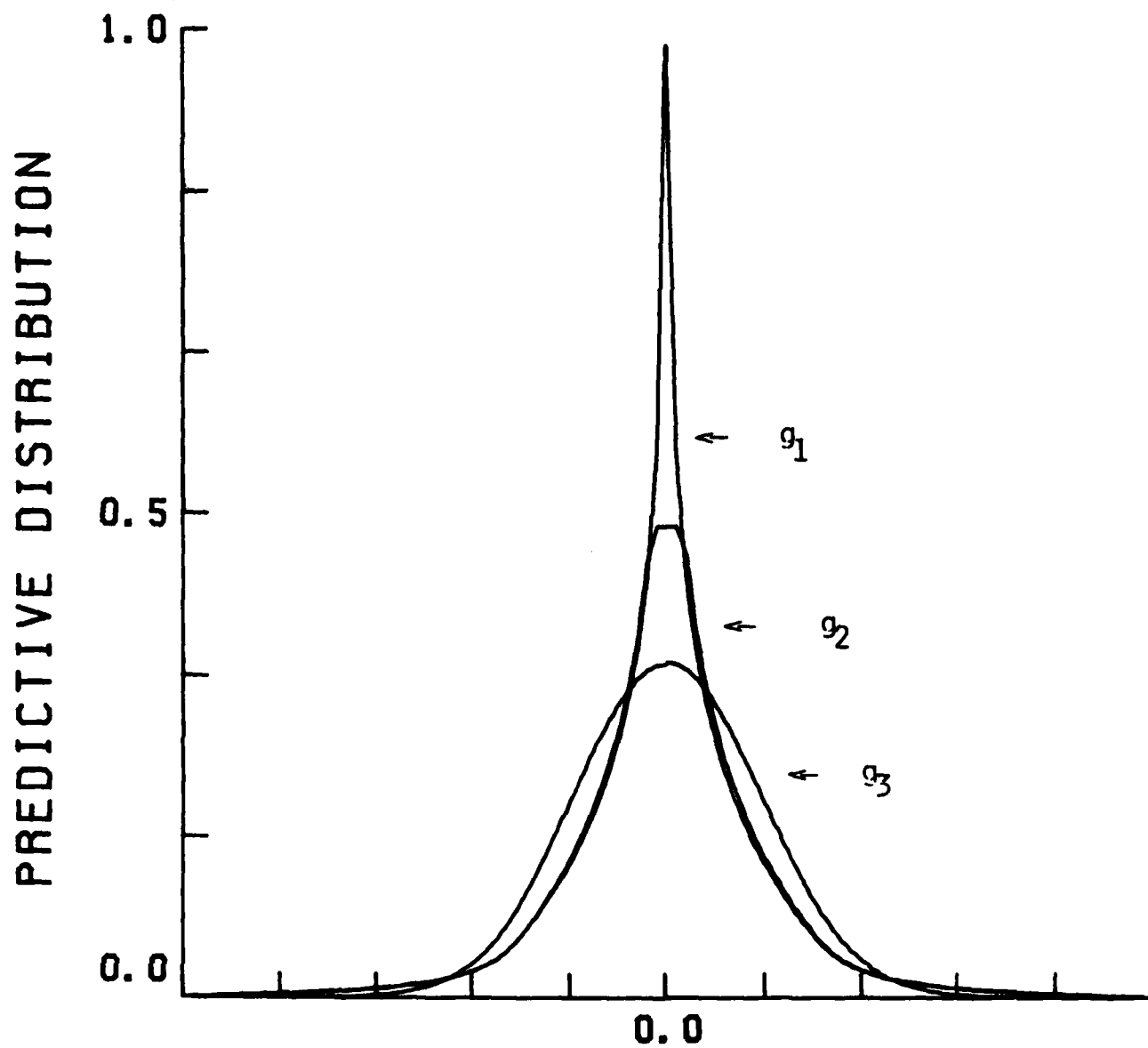
Fig. 1
Brockett

Fig. 2
Brockett

Fig. 3
Brockett

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CCS 467	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Information-Theoretic Non-Parametric Unimodal Density Estimation		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. Brockett, A. Charnes, K. Paick		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0413 N00014-82-K-0295
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies The University of Texas at Austin Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 434) Washington, D.C.		12. REPORT DATE February 1984
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

This document has been approved for public release and sale; its distribution is unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Information theory, M.D.I. density estimation, maximum entropy, unimodality, non-parametric density estimation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

We present an information-theoretic method for nonparametric density estimation which guarantees that the resulting density is unimodal. The method inputs data in the form of moment or quantile information and consequently can handle both data derived and non-data derived information. In the non-data derived situation it yields a method for obtaining unimodal Bayesian prior distribution.

DD ,FORM 1473 EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-014-6601 |

END

FILMED

DTIC